DOCUMENT RESUME

ED 292 113 CS 211 090

AUTHOR Iaccino, James F.; Spirek, Pamela
TITLE Pictorial Retention of Plausible vs. Bizarre Images as a Function of Unexpected Recall.
PUB DATE 88
NOTE 20p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Encoding (Psychology); *Figural Aftereffects; Long Term Memory; *Perception Tests; Pictorial Stimuli; *Recall (Psychology); Recognition (Psychology); *Retention (Psychology); Schemata (Cognition); Short Term Memory; Visualization; *Visual Learning; Visual Measures; Visual Perception; *Visual Stimuli
IDENTIFIERS Visual Response System

ABSTRACT

Previous research has demonstrated that bizarre imagery facilitates long-term recall of noun pairs. A study investigated the effects of bizarreness when more pronounced delays were used. Subjects, 40 introductory psychology students from Illinois Benedictine College, were shown 30 plausible and 30 bizarre scenes at a viewing distance of approximately 380 mm for a period of 10 seconds, asked to rate the pictures on their own image scale, and then given an unexpected recall test, one or two weeks later, based on the pictures viewed. Results demonstrated that bizarre noun pairs were more likely to be recalled than commonplace items with lengthier delays, lending support to the stimulus distinctiveness theory. Results also indicated that bizarre images aided recall only under the extended delay interval. (A table of data and a figure are included.) (JK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED292113

Pictorial Retention of Plausible vs. Bizarre Images as a Function of Unexpected Recall

James Iaccino and Pamela Spirek

Illinois Benedictine College

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

James Iaccino

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Running head: Pictorial Retention of Plausible vs. Bizarre Images

CS211090

Abstract

Past research has demonstrated that bizarre imagery facilitates long-term recall of noun pairs. An experiment was executed to investigate the effects of bizarreness when more pronounced delays were used. Undergraduate subjects were shown 30 plausible and 30 bizarre scenes, asked to rate the pictures on their own image scale, and were then given an unexpected recall test one- or two-weeks later. Results indicated that bizarre images aided recall only under the extended delay interval. We have suggested that interference over time may be less with bizarre images, especially when an incidental learning paradigm is employed.

Pictorial Retention of Plausible vs. Bizarre Images as a Function of Unexpected Recall

Does bizarre imagery facilitate retention? For centuries, the answer to this question was a definite yes. Unusual, strange and implausible relations between stimulus elements have been suggested to aid in retrieval of those items.

Two theories have been proposed to explain facilitation with bizarre items. The first concerns itself with subject motivation or interest. Bower(1970, 1972) raises the possibility that bizarreness might enhance the subject's motivation for learning, making the task more interesting. The novel associations might arouse and maintain interest that more normal pairings would not do. However, the basic problem with this theory is that motivation is difficult to measure empirically.

An alternative approach that has been referred to in current literature (O'Brien & Wolford, 1982) is the stimulus distinctiveness theory. Bizarre imagery might cause the relevant associations to become more salient, thus contributing to less confusion among the stimulus items, less interference over extended periods of time, and eventually more accurate recall.

In terms of empirical support, the findings have been contradictory. Some researchers found that bizarre noun pairs did not exert a significant effect on immediate recall; in fact, the normal commonplace scenes aided memory retrieval (Collyer, Jonides & Bevan, 1972; Nappe & Wollen, 1973; Wollen, Weber & Lowry, 1972). However, when unexpected recall was delayed (periods ranged from 24 hours to 15 weeks), bizarre images did help (Andreoff & Yarmey, 1976; Delin, 1968; Merry, 1980; Merry & Graham, 1978; O'Brien & Wolford, 1982). It would appear that under long-term recall conditions, the bizarre associations were retained to a greater degree than the normal relations, indicating that the very nature of bizarreness is to produce less interference over time. This would seem to tentatively support the stimulus distinctiveness theory. However, the lengths of the delay intervals and stimulus materials have varied from study to study so that an accurate interpretation is hard to arrive at.

The procedures by which bizarre materials have been presented have added their impact to the confusing results. No one methodology has been used consistently. Some researchers have requested that the subjects employ their own bizarre images to connect noun pairs (Collyer, et al., 1972; Delin, 1968; Nappe & Wollen, 1973; Wood, 1967). Others have used bizarre pictures or sentences provided

by the experimenter so that the images would be more standardized for the subject and could be more easily represented in the subject's memory (Andreoff & Yarmey, 1976; Merry, 1980; Merry & Graham, 1978). However, this methodology had its limitations as well. The criticism raised was that the experimenter images might not have the same "idiosyncratic strength" as those formed by the subjects themselves.

A new format developed by O'Brien and Wolford (1982) has attempted to more effectively control for bizarreness. First, judges rated pairs of words (that were high in imagery value) on a seven-point scale for plausibility or bizarreness. From these pairs, a select number having high inter-rater reliability were used in the main experiment. Further, subjects were asked to rate the words themselves (in terms of either plausibility or bizarreness) before being given a recall test. The subjects' ability to represent the words on their own scale, however, was not part of the data analysis. In our opinion, having both judge and subject rate the items should contribute to a more effective design and should be able to address itself to the main issue of experimenter vs. subject discrepancy in the use of bizarre imagery.

For the aforementioned reasons, the latter methodology was not only replicated but improved upon, by comparing subjects'

ratings of normalcy and bizarreness with the judges. Also, delay intervals ranging from one- to two-week periods were employed. Such time periods typically were not used in past research; even the O'Brien and Wolford study (1982) did not have longer delays past seven days, the rationale being that substantial loss in information would occur with greater interval periods, making data analysis ineffective. But, if bizarre imagery was suggested to improve recall performance the longer the delay, then it was hypothesized that the lengthier two-week period would make the bizarre effects more pronounced, even when poorer recall of items was anticipated. Finally, the typical undergraduate population of subjects was tested as was the case with most earlier studies, with the exception of Merry and Graham (1978) who examined 12- and 13-year-olds' recall performance.

Method

Subjects. Forty male and female introductory psychology students from Illinois Benedictine College at Lisle participated in the study. They were informed before the experiment began that they would receive extra-credit points in their psychology classes for volunteering their services.

Material and design. The materials consisted of 96 high-imagery nouns, ranging from 6.03 to 6.95 on the I scale of the

Paivio, Yuille and Madigan (1968) word lists. These nouns were also high in concreteness, showing a direct referencing to sensorial experiences, with the range extending from 6.20 to 7.00 on the corresponding C scale. Typical noun pairs had been obtained from the Paivio et al. lists (1968) in past research to control for image-evoking effects with bizarreness. The 96 items selected were arranged to form 48 pairs of words. For each pair, pictures were drawn and shaded in pencil on five by seven inch index cards; the drawings depicted the elements in both a bizarre and plausible interaction.

Then, three judges rated the line drawings on a seven-point scale, extending from most normal (1) to most bizarre (7). Data from these scales indicated high inter-rater reliability: coefficients of .92, .87, and .87 were obtained, suggesting a very high degree of consistency among the three judges. Mean ratings were 1.5 for the normal pairs and 6.1 for the more bizarre images.

From the 48 pairs, 30 were selected and used in the main study.

Insert Table 1 about here

These nouns elicited the highest degree of inter-rater agreement in terms of their normalcy and bizarreness. Again, each of the 30

pairs illustrated both a normal and unusual scene. The type of design employed was a two by two factorial combination of image type (plausible vs. bizarre) and delay of recall test (one-week vs. two-week), with type of image being a within-subjects factor.

Procedure. Subjects were informed previous to the start of the experiment that a psychology professor was interested in finding some common properties of bizarre images that would aid him/her in devising a new psychological test for bizarreness. They were then told that the task involved viewing a series of index cards and deciding whether each picture depicted a plausible (i.e., normal, commonplace) or bizarre (i.e., unusual, strange) interaction. Examples were given to ensure that subjects understood what was expected of them.

In the preliminary exercise, six noun pairs were presented (three normal and three bizarre scenes). On each card was the picture as well as the stimulus-response pair printed directly below the drawing. Subjects were instructed to look at each of the pictures for a period of 10 seconds, and then indicate on a sheet numbered 1 to 6 if the picture depicted a bizarre or normal scene. They simply wrote "N" for normal and "B" for bizarre on the rating sheet provided.

The main study began after subjects became acquainted with this rating system. Sixty cards were shown at a viewing distance of approximately 380 mm. Again, the same format was employed: each card was shown for a period of 10 seconds, after which time the subject wrote down (on a new sheet numbered 1 to 60) his/her rating. During this phase, if subjects inquired about the stimulus-response words pencilled in under the pictures, the experimenter explained that the words corresponded to those particular objects depicted in the pictures and that they were necessary to differentiate one scene from another. Problems were not anticipated provided this rationale, and in fact, difficulties did not occur.

After each subject viewed and categorized the 30 plausible and 30 bizarre images, s/he was asked to participate in an additional phase of the study. Each was asked to return to the same setting one- or two-weeks later, at which time a similar rating task would be administered. No mention of a recall task was made, so subjects were led to believe they would only categorize another deck of cards.

Upon returning, an unexpected recall test was given based on the pictures viewed previously. Most subjects (approximately four-fifths) did come back to participate in this phase; those that did not return were excluded from the study and their ratings were

omitted from the data analysis. The recall test for both the one- and two-week delay groups (ns = 20) consisted of presenting cards with either the stimulus or response word printed directly in the center. The half of each noun pair was randomly determined by the experimenter. The subjects' assignment was to write the corresponding member of each pair on an appropriately-numbered sheet of paper within a set time limit of 10 seconds. After the recall test was over, the true purpose of the research was revealed.

Results

In determining a correct response, the right word had to be written down. Synonyms or approximations were considered to be incorrect.

The data were analyzed in a two by two analysis of variance, with type of imagery (normal vs. bizarre) treated as a within-subjects variable and delay of test (one-week vs. two-weeks) treated as a between-subjects variable. As hypothesized, bizarre images were recalled better than plausible images, F (1,36) = 7.96, $p < .01$. Also, recall deteriorated significantly as test delay increased from one- to two-weeks, F (1,36) = 13.51, $p < .001$.

The most important finding was the interaction between type of image and delay of test, F (1,36) = 4.82, $p < .05$.

Insert Figure 1 about here

As portrayed in Figure 1, the mean number of plausible items recalled dropped significantly from the first to the second week. Subsequent Newman-Keuls analyses revealed that the two-week palusible-item condition differed from the other conditions. Moreoever, bizarre noun pairs apparently were recalled at the same level regardless of the delay interval.

In addressing the issue of subject vs. experimenter discrepancy in rating images as either plausible or bizarre, rank-order correlations were performed comparing subjects' ratings of the items with the judges' ratings. Significant correlations indicated a high degree of agreement for both bizarre and plausible images, respectively $r_{BB} = .37$, $p < .05$ and $r_{NN} = .62$, $p < .001$.

Discussion

It was predicted that bizarre images would facilitate recall at both one- and two-week intervals, with effects being strongest at two weeks. Results did demonstrate that bizarre noun pairs were more likely to be recalled than commonplace items with lengthier delays, lending support to the stimulus distinctiveness theory

(O'Brien & Wolford, 1982). Interference over long periods of time may be less with bizarre than plausible images, thus explaining their better retention.

However, under the one-week delay condition, recall performance with the bizarre pairs did not differ significantly from the more normal items. These results are inconsistent with the current experimental prediction as well as previous methodologies that have found bizarre effects using similar delay intervals (Merry, 1980; Merry & Graham, 1978; O'Brien & Wolford, 1982). In fact, the latter researchers had shown that a delay of just five days would enhance the effects of bizarre imagery over commonplace items.

Even though this study was a replication of the O'Brien and Wolford procedure, several differences should be noted which might account for the discrepant results. First, more noun pairs were used, 60 compared to the previous 48 pictures. Possibly the number of items to be recalled might interact with bizarreness, with a smaller memory load enhancing bizarre image recall sooner.

Second, and more importantly, the type of task was deliberately executed as an incidental recall test. Subjects were specifically informed to ignore the stimulus-response words printed directly underneath the pictures. In the O'Brien and Wolford study (1982), these instructions were not clearly expressed; though no indication

of a recall test was made at the time of image categorization, subjects still could have been intentionally learning the word pairs. They might even have been anticipating a test on the retention of these pairs and so focused their concentration on the more unusual associations, forming more durable traces of these bizarre items in memory. Thus, whether the recall test was expected or not might affect retention level of particular items, especially bizarre figures.

Finally, subjects demonstrated a high level of agreement with the judges in rating the current items either bizarre or commonplace. Subjects' ratings were totally ignored in the O'Brien and Wolford methodology. It is very probable that the items the previous judges rated as unusual were not those that subjects regarded as bizarre, severely questioning the interpretation of these researchers' results. Therefore, the number of items to be recalled, the type of recall task used (whether an intentional or incidental paradigm) and a measure of subject-judge consistency in rating the items are all variables that need to be considered when designing studies and deriving conclusions with bizarre images.

Future studies should definitely investigate the possibility of studying long-term retention with bizarreness at more pronounced

durations to further support the stimulus distinctiveness theory. Implications for educational and instructional techniques should also be considered.

References

Andreoff, G. R. & Yarmey, A. D. Bizarre imagery and associative learning: a confirmation. Perceptual and Motor Skills, 1976, 43, 143-148.

Bower, G. Analysis of mnemonic device. American Scientist, 1970, 58, 496-510.

Bower, G. Mental imagery and associative learning. In L. Gregg (Ed.), Cognition in learning and memory. New York: John Wiley, 1972.

Collyer, S. C.; Jonides, J.; & Bevan, W. Images as memory aids: is bizarreness helpful? American Journal of Psychology, 1972, 85, 31-38.

Delin, P. S. Success in recall as a function of success in implementation of mnemonic instructions. Psychonomic Science, 1968, 12, 153-154.

Merry, R. Image bizarreness in incidental learning. Psychological Reports, 1980, 46, 427-430.

Merry, R. & Graham, N. C. Image bizarreness in children's recall of sentences. British Journal of Psychology, 1978, 69, 315-321.

Nappe, G. W. & Wollen, K. A. Effects of instructions to form common and bizarre mental images on retention. Journal of Experimental Psychology, 1973, 100, 6-8.

O'Brien, E. J. & Wolford, C. R. Effect of delay in testing on retention of plausible versus bizarre mental images. Journal of Experimental Psychology: Learning, Memory and Cognition, 1982, 8, 148-152.

Paivio, A.; Yuille, J. C.; & Madigan, S. A. Concreteness, imagery and meaningfulness values for 925 nouns. Journal of Experimental Psychology Monograph Supplement, 1968, 76 (1, Pt. 2).

Wollen, K. A.; Weber, A.; & Lowry, D. H. Bizarreness versus interaction of mental images as determinants of learning. Cognitive Psychology, 1972, 3, 518-523.

Wood, G. Mnemonic systems in recall. Journal of Educational Psychology (Monograph), 1967, 58 (6, Pt. 2).

TABLE 1

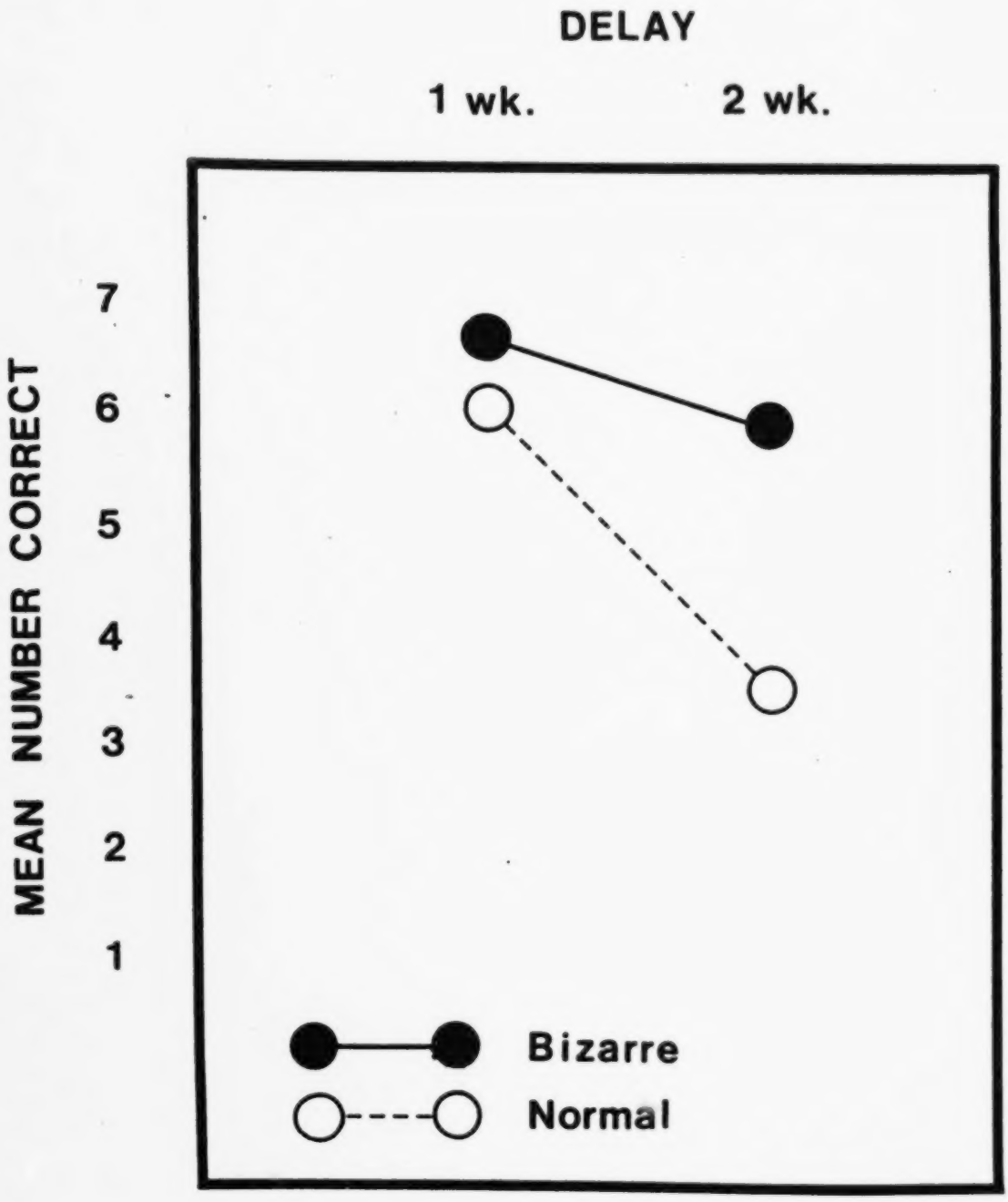
Stimulus-Response Pairs

Used with Image Type

1.	piano - cigar	16.	sea - ship
2.	oven - hammer	17.	bullet - revolver
3.	daffodil - table	18.	forehead - insect
4.	barrel - frog	19.	hairpin - girl
5.	skull - bottle	20.	pencil - boy
6.	elephant - shoes	21.	plank - house
7.	microscope - lobster	22.	tomb - flower
8.	tank - tower	23.	earth - volcano
9.	clock - wine	24.	bird - book
10.	missile - star	25.	lemon - toast
11.	doctor - bowl	26.	dollar - fork
12.	cat - chair	27.	fur - glacier
13.	apple - fox	28.	lake - alligator
14.	butterfly - car	29.	truck - engine
15.	storm - umbrella	30.	stagecoach - strawberry

Figure Caption

Figure 1. Mean number of correct responses by image type and by delay.


DELAY
1 wk.
2 wk.
MEAN NUMBER CORRECT
7
6
5
4
3
2
1
Bizarre
Normal